Heredity

REVIEW
Molecular cytogenetic and genomic insights into chromosomal evolution

A Ruiz-Herrera[1,2], M Farré[1] and TJ Robinson[3]

This review summarizes aspects of the extensive literature on the patterns and processes underpinning chromosomal evolution in vertebrates and especially placental mammals. It highlights the growing synergy between molecular cytogenetics and comparative genomics, particularly with respect to fully or partially sequenced genomes, and provides novel insights into changes in chromosome number and structure across deep division of the vertebrate tree of life. The examination of basal numbers in the deeper branches of the vertebrate tree suggest a haploid (n) chromosome number of 10–13 in an ancestral vertebrate, with modest increases in tetrapods and amniotes most probably by chromosomal fissioning. Information drawn largely from cross-species chromosome painting in the data-dense Placentalia permits the confident reconstruction of an ancestral karyotype comprising $n=23$ chromosomes that is similarly retained in Boreoeutheria. Using in silico genome-wide scans that include the newly released frog genome we show that of the nine ancient syntenies detected in conserved karyotypes of extant placentals (thought likely to reflect the structure of ancestral chromosomes), the human syntenic segmental associations 3p/21, 4pq/8p, 7a/16p, 14/15, 12qt/22q and 12pq/22qt predate the divergence of tetrapods. These findings underscore the enhanced quality of ancestral reconstructions based on the integrative molecular cytogenetic and comparative genomic approaches that collectively highlight a pattern of conserved syntenic associations that extends back ~360 million years ago.
Heredity (2012) 108, 28–36; doi:10.1038/hdy.2011.102; published online 23 November 2011

INTRODUCTION

How genomes are organized and which types of chromosomal rearrangements are implicated in speciation and macroevolutionary events are fundamental to understanding the dynamics of chromosomal evolution. Molecular cytogenetic data and the increasing availability of partially or fully sequenced genomes from a variety of vertebrate species have fueled advances in phylogenomics (phylogenetic reconstructions using genomic data). This has lead to hypothesized ancestral chromosome numbers, karyotypes and the identification of conserved chromosomal syntenies and segmental associations at different taxonomic levels.

Chromosome number variation has traditionally been considered a proxy for the structural modification of karyotypes, especially so in groups of organisms where detailed information such as the differential staining of chromosomes, the extent and location of heterochromatin, and the location and number of nucleolar organizers is lacking. A considerable body of early work on chromosome number variation was reviewed by White (1973), who expressed reservations on whether it would be possible to determine 'modal numbers' for groups of organisms the higher one progresses in the systematic hierarchy. More specifically, he was of the view that 'to speak of a type number for the Insecta, the Vertebrata or even the Mammalia would be absurd'. However, recently, various computational approaches have been used to estimate the extent of rearrangement

events and to derive the putative genomic architecture of ancestral genomes by inferring evolutionary histories from entire genomes. This has led to suggestions of ancestral syntenies and chromosomal complements—each progressively more distant in divergence—for amniotes (~310 million years ago, mya), tetrapods (~360 mya) and even vertebrates (~450 mya) (Postlethwait et al., 2000; Naruse et al., 2004; Woods et al., 2005; Kohn et al., 2006; Nakatani et al., 2007; Voss et al., 2011).

Among vertebrates, phylogenomic investigations have focused principally on mammalian genome evolution, in large part reflecting the availability of chromosomal and genomic information for this clade. Extant mammals (represented by monotremes, marsupials and placental or eutherian mammals) last shared common ancestry nearly 162 mya (Hallström and Janke, 2010). Modern eutherian taxonomic schemes recognize four superordinal clades (Afrotheria, Xenarthra, Laurasiatheria and Euarchontoglires) largely on the basis of phylogenetic analysis of both nuclear and mitochondrial DNA (Hallström and Janke, 2010 and references therein) and insertion sites of retroelements (Nishihara et al., 2005; Kriegs et al., 2006; Waters et al., 2007; Churakov et al., 2009). Although it would appear that the terms 'Eutherian' and 'Boreoeutherian' have been used synonymously in comparative cytogenetic and phylogenomic studies, they do in fact represent different nodes. Eutheria refers to everything on the so-called 'eutherian' side of the 'metatherian'-'eutherian' dichotomy

[1]Departament de Biologia Cellular, Fisiologia i Immunologia, Universitat Autònoma de Barcelona, Barcelona, Spain; [2]Institut de Biotecnologia i Biomedicina, Universitat Autònoma de Barcelona, Barcelona, Spain and [3]Evolutionary Genomics Group, Department of Botany and Zoology, University of Stellenbosch, Matieland, South Africa
Correspondence: Professor TJ Robinson, Evolutionary Genomics Group, Department of Botany and Zoology, University of Stellenbosch, Private Bag X1, Stellenbosch, Matieland 7602, South Africa.
E-mail: tjr@sun.ac.za

(that is, Afrotheria, Xenarthra, and Boreoeutheria, and all fossil relatives that are more closely related to this clade than to Marsupialia). Boreoeutheria on the other hand comprises Laurasiatheria (Waddell et al., 1999) and Euarchontoglires (Murphy et al., 2001a). It is also more accurate to refer to the 'eutherian' ancestral karyotype as that of Placentalia as the data we have to infer this from are solely from placentals (all extant members of the last common ancestor to Atlantogenata (Afrotheria+Xenarthra) and Boreoeutheria; Asher and Helgen, (2010))—a usage that we follow in this review.

Here we examine how molecular cytogenetics and the in silico analysis of genomic sequences have contributed to our understanding of mammalian chromosomal evolution and the identification of conserved genomic regions. Furthermore, we review and extend previous observations by providing new data on the presence of conserved syntenic segmental associations that track back to the origin of tetrapods.

CHROMOSOME NUMBER VARIATION IN VERTEBRATES

Chromosome number and the number of chromosomal arms are good summary statistics of karyotypic change and hence chromosomal evolution in groups of organisms. Although data on chromosome arm number variation (the nombre fondamental of Matthey (1945), usually abbreviated to NF) are sparse, information on chromosome numbers across high-level taxonomic groups (mammals, birds, reptiles and amphibians) is extensive (Figure 1). Early lists of animal haploid (n) or diploid (2n) numbers (reviewed by White, 1973) included those on insects, crustaceans, fishes and, with respect to mammals, those of Hayman and Martin (1969) for marsupials and Matthey (1958) for placentals. Although these early attempts often suffered from poor taxon representation, they nonetheless permitted several general conclusions one of which was that the haploid number of most animal species lies between 6 and 24.

Since these early investigations, the biggest advances in determining chromosome numbers in deep branches of the vertebrate tree of life have, not surprisingly, resulted from the in silico scans of sequenced genomes. Genomic comparisons between human and teleost fish species (medaka, zebrafish and tetraodon) permitted hypothesized ancestral vertebrate genome configurations with n=10–13 (Postlethwait et al., 2000; Naruse et al., 2004; Woods et al., 2005). Detailed analyses of the likely amniote and tetrapod compositions followed (Khon et al., 2006; Nakatani et al., 2007). These studies, based on data from human, chicken, zebrafish, medaka and pufferfish genomes (Khon et al., 2006), were subsequently expanded (Nakatani et al., 2007) to include the
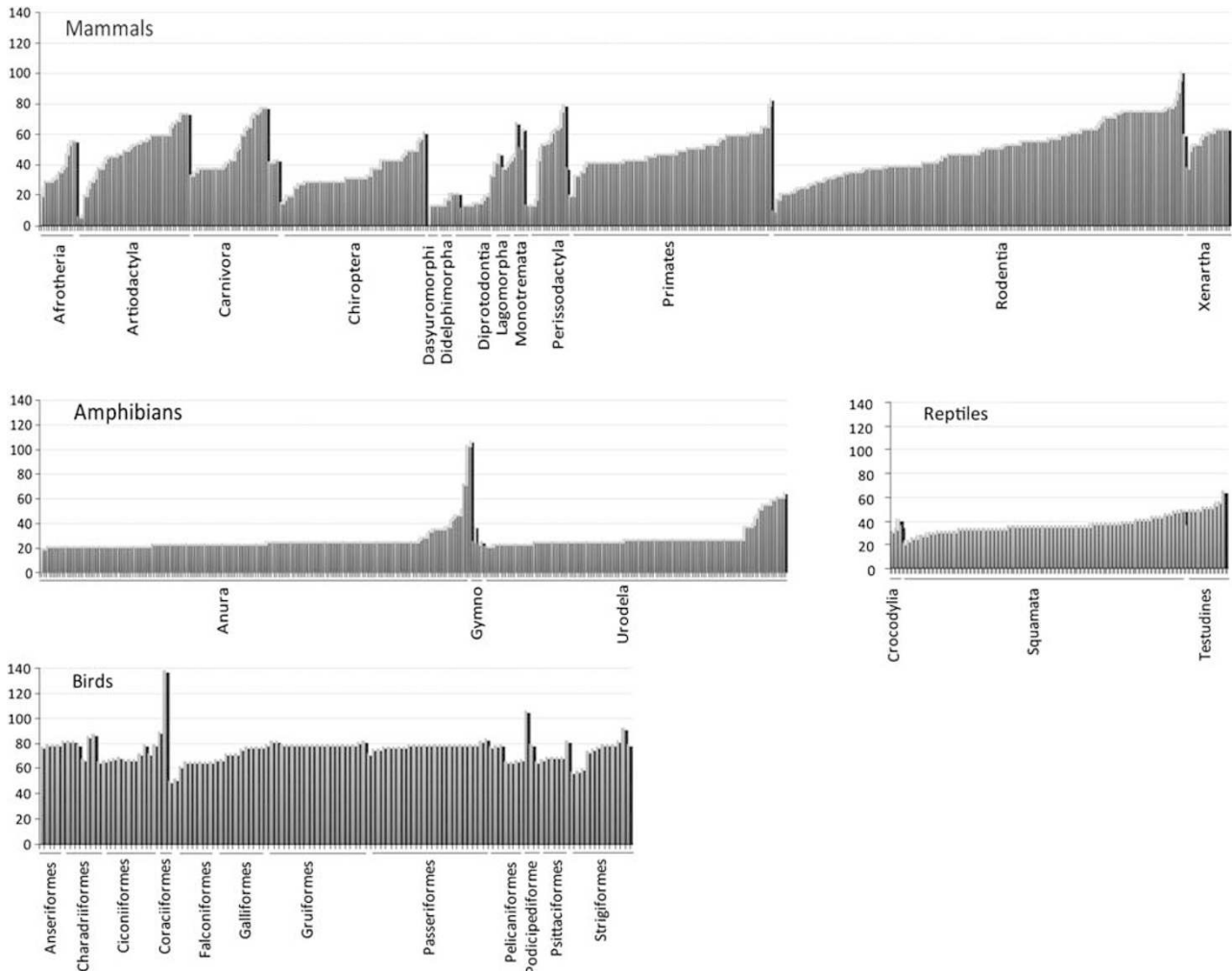

Figure 1 Chromosomal number variation among vertebrates. The x axis indicates the diploid chromosomal number, whereas the y axis groups species in different orders. The data for each taxonomic group are based on 515 species of mammals, 117 species of birds, 170 reptiles and 328 amphibians. Chromosomal data extracted from O'Brien et al. (2006) and Gregory (2011). A full color version of this figure is available at the Heredity journal online.

mouse, dog, tunicate (*Ciona intestinalis*) and sea urchin (*Strongylocentrotus purpuratus*). Khon et al. (2006) relied on 'E-painting'—the *in silico* identification of orthologous gene pairs to identify conserved genomic regions—whereas Nakatani et al. (2007) developed their own computational methodology to detect 'Ohnologs' (paralogs produced by two rounds of whole genome duplication) and thus conserved vertebrate linkage blocks. These studies, respectively, posit $n=18$ for the tetrapod ancestor (using the teleost pufferfish as outgroup), and $n=26$ for the amniote ancestor (using the pufferfish and medaka as outgroups).

The basal numbers retrieved by the various studies outlined above collectively permit inferences on the broader patterns of chromosome number changes across these groups. First, the low chromosome number suggested for the tetrapod ancestor increased to 26 in the amniote ancestor, most probably by multiple fissions. Previous attempts to reconstruct the ancestral tetrapod genome configuration (Kohn et al., 2006; Voss et al., 2011) have resulted in contradictory outcomes. Kohn et al., 2006 proposed an ancestral tetrapod karyotype with $n=18$. On the other hand Voss et al. (2011), who studied the *Xenopus* ($n=10$) and *Ambystoma* ($n=14$) as part of an investigation into ancestral tetrapod chromosomes, proposed a high but unspecified chromosome number mirroring those usually found in birds. Their hypothesis is based on the observation that phylogenetically derived lineages (such as *Xenopus* and *Ambystoma*) have fewer chromosomes, indicating a tendency to have reduced chromosome numbers in these lineages. In contrast, birds (represented here by chicken) and mammals (by platypus, opossum and 11 placental species) are characterized by markedly different modes of chromosome number evolution.

Aves

The predominant mode of genome reorganization in Aves is chromosomal fission. Avian karyotypes are composed of microchromosomes and macrochromosomes but contrary to non-avian reptiles, birds are characterized by high chromosomal numbers that range from $n=20$ (or 21; see Nie et al., 2009) to $n=69$ (De Smet, 1981; Figure 1). Descriptions of the ancestral avian karyotype are conventionally based only on macrochromosomes (Griffin et al., 2007; Nanda et al., 2011) and suggest that many of these have remained conserved within the group without disruption by interchromosomal rearrangements (reviewed in Ellegren, 2010). In fact, Griffin et al. (2007) have argued that the ancestral avian karyotype was similar to that of chicken, with macrochromosomes 1, 2, 3, 4q, 5, 6, 7, 8, 9, 4p and Z representing the ancestral state for chromosomes 1–10+Z; chromosome 4 was regarded as the most ancient linkage group within this karyotype.

Mammalia

A different situation holds for Mammalia where significant variation in chromosomal number is observed among Monotremata, Marsupialia and the eutherian placental mammals (Placentalia; Figure 1).

The three extant species belonging to Monotremata all have high diploid chromosome numbers with platypus characterized by $n=26$, and both the short-beaked and long-beaked echidnas having $n=32$ (O'Brien et al., 2006). Although only one of these species was included in our analysis (the platypus, whose genome has been sequenced and is partially assembled), it is nonetheless clear that, as with Aves, fission events predominate in the karyotypic evolution of Monotrema.

Comprehensive cytogenetic studies on marsupials show that chromosomal numbers within the group range from $n=5$ to $n=16$ (Hayman, 1990). Whereas the majority of the families have conserved karyotypes (mainly $n=7$), the Macropodidae (kangaroos, wallabies and rat-kangaroos) shows evidence of more extreme chromosome

reshuffling including fusion/fissions, inversions and centromere repositioning (O'Neill et al., 2004 and references therein). Among marsupials, the South American opossum (*Monodelphis domestica*) is the only marsupial for which pair-wise alignments with the human genome are possible. Recently Westerman et al. (2010), using a combination of cytogenetics and sequence-based phylogenetics, have argued that the karyotype of the opossum ($n=9$) is highly conserved in relation to those of Australian marsupials confirming previous hypotheses (Rens et al., 2001). *Monodelphis domestica* groups within the basal Didelphimorphia (Nilsson et al., 2010; Westerman et al., 2010) and is thought to have undergone two fissions from the hypothesized marsupial ancestral karyotype of $n=7$ (Rens et al., 2001). If the marsupial ancestral estimate is correct (our small sample size precludes an estimate for Marsupialia given that only one fully sequenced genome is available), a dramatic decrease in chromosome number appears to have occurred in the marsupial lineage (presumably by serial fusion events) since its divergence from the mammalian common ancestor (with $n=23$—see mammalian ancestral configuration discussed below) ~138 mya (Hallström and Janke, 2010).

The extremes in mammalian chromosome number occur in the species-rich Placentalia where these range from $n=3$ in the female Indian muntjac to a high of $n=51$ in the Red viscacha rat (O'Brien et al., 2006). There is also substantial variation among Orders (Figure 1) reflecting the complex dynamics of mammalian chromosomal evolution. Recent studies based on cross-species chromosome painting analyses have estimated an ancestral haploid chromosome number that ranges from 22 to 25 for Placentalia (Chowdhary et al., 1998; Froenicke et al., 2003; Richard et al., 2003; Yang et al., 2003; Svartman et al., 2004, 2006; Murphy et al., 2005; Ferguson-Smith and Trifonov, 2007), with a consensus opinion settling on $n=23$ (see Ferguson-Smith and Trifonov, 2007). The rationale underpinning this, and the likely composition and uniqueness of the ancestral karyotype, as well as its correspondence with *in silico*-based studies of genome sequences, are discussed below.

ANCESTRAL PLACENTAL KARYOTYPES AND THE DETECTION OF SYNTENIES BASED ON FISH

Reconstructions of ancestral karyotypes across the placental mammalian tree rely heavily on molecular cytogenetic approaches that entail cross-species fluorescence *in situ* hybridization (Zoo-FISH; methodology reviewed by Rens et al., 2006) using human and chromosome-specific DNA sequences from other species as probes. This has allowed the identification of orthologous regions defined by their correspondence with human chromosomes, and the delimitation of chromosomal rearrangements among species. These conserved regions span entire chromosomes, chromosomal arms, or chromosomal segments in closely and distantly related placental species permitting the generation of large-scale comparative maps among taxa. In the present context, it is important to make the distinction between segmental associations (the adjacent syntenies of some terminologies) and syntenic blocks that are retained *in toto* among lineages. The detection of segmental associations such as 4q/8p/4pq, 3p/21, 14/15, 10p/12pq/22qt, 16q/19q, 7a/16p and 12qt/22q (each of which involve segments of human chromosomes that in combination correspond to complete chromosomes in the ancestral eutherian karyotype) in placentals, chicken and opossum was based on the evidence of the entire adjacent segment having been retained in representative genomes (Robinson and Ruiz-Herrera, 2008). However, the incomplete nature of the genome assemblies of platypus and frog does not permit the same level of resolution. We consequently used the junction as the defining character of a particular conserved segmental association based on the

premise that the independent assembly of a precisely shared association in different lineages was unlikely. Gene order within the abutting syntenic blocks may be altered by intrachromosomal rearrangement, and the size of these segments affected by subsequent translocation of parts to other regions of the genome (Robinson and Seiffert, 2004).

In most high-level reconstructions the identification of conserved syntenic blocks in multiple extant species (that is, commonality) was taken to reflect the retention of a shared ancestral evolutionary state leading to hypothesized ancestral karyotypes for Placentalia (Chowdhary et al., 1998; Richard et al., 2003; Yang et al., 2003; Svartman et al., 2004, 2006; Ferguson-Smith and Trifonov, 2007) and various orders of mammals, principally within Boreoeutheria. Reconstructions of the placental ancestral karyotype (PAK) have diploid numbers that vary from $n=22$ to $n=25$ (see Table 1 in Svartman et al., 2004). The differences in interpretation are primarily related to the recognition of a single large chromosome (corresponding to HSA 1) in the placental ancestor (Murphy et al., 2003), the detection of the 10q/12p/22q conserved syntenic segmental association (Froenicke et al., 2003), and fusion of HSA1/19p (Yang et al., 2003, $n=22$) based on its presence in Afrotheria (aardvark, elephant, golden mole and elephant shrew), at the time regarded as the most basal split in the eutherian tree (Murphy et al., 2001a, b). The more recent studies appear, however, to have converged on $n=23$ for Placentalia (that is, the eutherian ancestral karyotypes of Froenicke et al., 2003; Wienberg, 2004; Ferguson-Smith and Trifonov, 2007), and an identical $n=23$ in the boreoeutherian ancestral karyotype (BAK; Froenicke, 2005; Froenicke et al., 2006; Robinson et al., 2006).

The most definitive of the PAK constructs (Figure 2a) benefited from the availability of genome sequence information from two important outgroup species, the opossum and chicken. This permitted the distinction between shared ancestral characters (symplesiomorphies) and those that are unique to the ingroup Placentalia (that is, showing shared derived similarity and referred to as synapomorphies) allowing firm conclusions on the evolutionary history of each (Robinson and Ruiz-Herrera, 2008). The PAK is considered to comprise two chromosome pairs (corresponding to human chromosomes 13 and 18) and three conserved chromosome segments (10q, 19p and 8q in the human karyotype) that are probable symplesio-

morphies as they are also present as unaltered orthologues in one or both outgroup species. Seven additional syntenic segmental associations (4q/8p/4pq, 3p/21, 14/15, 10p/12pq/22qt, 16q/19q, 7a/16p and 12qt/22q), each involving human chromosomal segments from two or more human chromosomes, are also present in one or both outgroup taxa and are probable symplesiomorphies. Importantly, however, there are eight intact pairs (corresponding to human chromosomes 1, 5, 6, 9, 11, 17, 20 and the X) and three chromosomal segments (7b, 2p-q13 and 2q13-qter) that are derived characters, potentially consistent with placental monophyly. In summary therefore, the karyotype of the putative ancestor of Placentalia comprised 32 conserved segments (including the X) and nine syntenic segmental associations, several of which trace back to a common amniote ancestor (discussed below; Figures 2a and b).

There is, at this point, no evidence to suggest that the boreoeutherian ancestral karyotype (Froenicke et al., 2006; Robinson et al., 2006) underwent further modification from the hypothesized PAK (see above). The subsequent radiation of Boreoeutheria, however, showed extensive karyotypic modification in most lineages permitting hypothesized ancestral karyotypes for several orders of mammals, as well as the identification of syntenic segmental associations that underpin the monophyly of various supraordinal and ordinal groups (Robinson et al., 2004; Wienberg, 2004; Froenicke, 2005; Ferguson-Smith and Trifonov, 2007; Ruiz-Herrera and Robinson, 2007, among others).

IN SILICO DETERMINATION OF THE ANCESTRAL BOREOEUTHERIAN KARYOTYPE AND EXTENT OF CONCORDANCE WITH THE CYTOGENETIC DATA

Advances from large-scale genome sequencing projects and the availability of new mathematical algorithms have revolutionized the study of chromosome evolution. The genomes of 35 mammalian species have been sequenced to differing degrees of completion (Ensembl database, version 59): 16 species of the Euarchontoglires (guinea pig, rat, mouse, rabbit, kangaroo rat, squirrel, tree shrew, tarsier, mouse lemur, bushbaby, marmoset, macaque, chimpanzee, orangutan, gorilla and human), 11 laurasiatherian representative (megabat, microbat, shrew, dolphin, pig, cow, alpaca, horse, dog, cat and hedgehog), three Afrotherian species (elephant, hyrax and tenrec), two xenathrans (sloth and armadillo), two species of Metatheria (wallaby and opossum) and the platypus as a prototherian representative. Of these, only the genomes of chimpanzee, rhesus macaque, orangutan, mouse, rat, cow, dog, horse and pig are sufficiently complete to allow pair-wise alignments with the human genome and the delimitation of syntenic blocks with a high degree of confidence.

Several sequenced-based reconstructions of the boreoeutherian ancestral karyotype have been attempted, often resulting in disparate outcomes compared with the findings suggested by FISH. In general terms, two different approaches can be distinguished when defining ancestral genomes in this way; (i) those that rely on the minimal number of rearrangements required to obtain the syntenies that lead to modern genomes (that is, MGR, Bourque and Pevzner, 2002) or (ii) models that focus on identifying conserved synteny blocks (Ma et al., 2006). The former methodology was used in an early attempt at the reconstruction of a mammalian (but more correctly boreoeutherian) ancestral karyotype (human–rat–mouse) using the chicken as an outgroup (Bourque et al., 2005). Although there is reasonable correspondence in the chromosome numbers suggested by MGR and cytogenetic data ($n=21$, cf. the $n=23$ posited by most chromosome painting strategies), the numbers of conserved segments and the numbers of syntenic associations were vastly different (Froenicke

Table 1 Number of orthologous genes and homologous synteny blocks in species established by pairwise comparisons to human

Species	No. of orthologs genes	No. of HSBs	Median length (bp)[a]	Genome representation (%)[b]
Chimpanzee	115 835	39	66 808 733	87.43
Orangutan	113 962	60	33 425 697	83.32
Macaque	117 410	72	29 853 043	91.37
Mouse	120 957	275	4 542 283	89.92
Rat	118 266	278	4 912 642	93.03
Cow	120 340	228	5 792 994	85.97
Dog	116 105	189	7 290 339	89.51
Armadillo	93 193	268	75 290	0.67
Elephant	113 657	141	6 285 988	54.66
Tenrec	100 672	580	78 347	1.64
Opossum	115 284	472	2 452 061	89.11
Platypus	102 547	957	116 845	37.15
Chicken	96 838	468	887 305	94.89
Frog	99 597	1128	224 864	37.95

Abbreviations: bp, base pair; HSBs, homologous synteny blocks.
[a]Median length of HSBs.
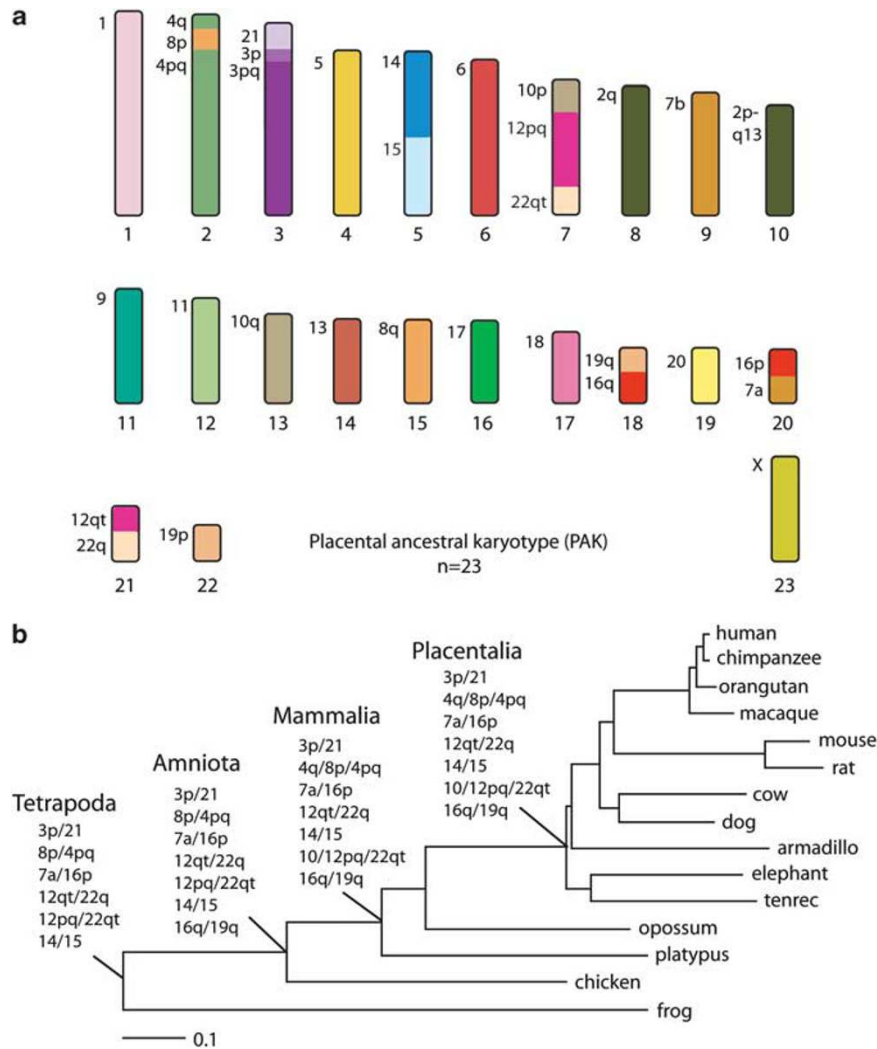[b]Percentage of each genome covered by our scans.

Figure 2 (**a**) Ancestral karyotype of Placentalia (PAK) defined by chromosomal correspondence to human chromosomes. Note that the HSA3/21 junction corresponds to human chromosomal segment 3p (depicted in violet), a region close to the centromere (from position 76.0 to 87.0 Mbp; Ruiz-Herrera and Robinson, 2007; Robinson and Ruiz-Herrera, 2008), and the conserved segmental association should more correctly be referred to as HSA3p/21. (**b**) Phylogenetic tree showing syntenic segmental associations detected at each ancestral node: p, short arm; pq, segment comprises parts of both the short and long arms; q, long arm; qt, terminal portion of the q arm.

et al., 2006). The MGR approach resulted in only four syntenic segmental associations (3/21, 4/8, 12a/22a and 12b/22b) being in common with those suggested by molecular cytogenetics. The degree of concordance was improved by Murphy *et al.* (2005), who used both genomic sequence data and information from radiation hybrid maps of eight species to obtain a more comprehensive view of the dynamics of genome organization in mammals. Their computational approach proposed an ancestral chromosome number of *n*=24 and showed that 80% of the conserved segments are in common with those detected by molecular cytogenetic approaches. However, only half of the syntenic segmental associations (specifically 3/21, 4/8×2, 7/16, 14/15, 12/22 and 16/19) were shared by both approaches (Robinson *et al.*, 2006). Although it could be argued that the difference in the numbers of conserved segments is a reflection of the increased discrimination of the DNA sequence comparisons, several of the *in silico* syntenies fall within the limits detectable by FISH leading Froenicke *et al.* (2006) to question the effectiveness of the computational methodology. Using a different approach, in this case inferring contiguous ancestral regions

within the completed genomes of human, dog, rat and mouse with chicken and opossum as outgroups, Ma *et al.* (2006) posit an ancestral boreoeutherian karyotype with *n*=29 but, importantly, with strong support for five of the ancestral syntenic segmental associations proposed by cytogenetic methods (4/8, 3/21, 14/15, 12/22×2).

Although there is consensus among the cytogenetic and computational approaches with respect to those conserved syntenies with strong probabilistic support (3/21, 4/8, 14/15, 12a/22a and 12b/22b), there are a meaningful number of ambiguous adjacent syntenies in conflict with the cytogenetic model (specifically 1/22, 5/19, 2/18, 1/10 and 2/20). This has led to the integration of available algorithms (Alekseyev and Pevzner, 2009) and to new methods of genome sequences analysis (Peng *et al.*, 2009; Lin *et al.*, 2010; Pham and Pevzner, 2010). It is anticipated that these efforts may, in future, provide more consistency to ancestral reconstructions based on *in silico* analysis and the degree of correspondence to the boreoeutherian construct suggested by the molecular cytogenetic analysis of more than 100 taxonomically diverse mammalian species.

IN SILICO IDENTIFICATION OF SYNTENIC SEGMENTAL ASSOCIATIONS AT DEEPER NODES OF THE VERTEBRATE TREE

As Zoo-FISH across the eutherian/metatherian boundary has been unsuccessful (with the exception of a small portion of the X that is conserved between the two lineages, Glas *et al.*, 1999), there is a reliance on *in silico* methodologies to define the vertebrate protokaryotype, and to detect ancestral chromosomal syntenies that have been retained over deep diversification nodes. The recent publication (Hellsten *et al.*, 2010) of the first amphibian genome to be sequenced—that of *Xenopus tropicalis*, a lineage that is thought to have diverged from amniotes ~360 mya—offers an opportunity to revisit putative ancestral karyotypes and conserved syntenies (which indicate the likely structure of ancestral chromosomes), deep within the vertebrate tree of life.

We used the SyntenyTracker (Donthu *et al.*, 2009) to establish homologous synteny blocks (HSBs) between human and the genomes of 12 mammalian species (chimpanzee, orangutan, rhesus macaque, mouse, rat, cow, dog, armadillo, elephant, tenrec, opossum and platypus) plus the chicken and the frog (see online appendix for details). Table 1 provides the number of genes analyzed for each species and the number of HSBs detected, whereas the composition of the HSBs in the three progressively distant taxa to Placentalia—opossum, platypus and the chicken—is presented in Figure 3. Unfortunately, the draft frog genome is not assembled into chromosomes at this stage thus precluding the analysis of the whole karyotype and limiting our ability to unambiguously distinguish homologous and homoplasious syntenic associations between very distantly related species with potentially highly rearranged genomes (of which only portions can be traced in HSBs). The same shortcoming applies to platypus where several chromosomes remain unassembled (see below). The *in silico* chromosomal homologies identified by this approach permitted testing for PAK ancestral syntenic segmental associations at different phylogenetic levels (Figure 2b). This also allowed us to revisit the ancestral amniote and tetrapods genome compositions suggested by Kohn *et al.* (2006) and Nakatani *et al.* (2007) using syntenies identified in the frog, chicken, platypus and opossum.

Mammalian ancestral configuration

Our scans of the opossum and chicken genomes, analyzed as part of attempts to define placental chromosomal characters that define the monophyly of the group (Robinson and Ruiz-Herrera, 2008), revealed syntenic segmental associations (4q/8p/4pq, 3p/21, 14/15, 10p/12pq/22qt, 16q/19q, 7a/16p and 12qt/22q) that are shared with either (or, in some instances, both) opossum and chicken. No conserved human chromosomal segments were observed in the assembled platypus chromosomes (most probably because of low coverage of the annotated sequences). Nevertheless, there were several contigs (orthologous regions of small size, not yet assembled) that contained some of the syntenic segmental associations considered to be present in the PAK (Figure 2b). These were 4q/8p (Ultracontig173), 12qter/22q (Ultracontig252), 7a/16p (Ultracontig371), 3p/21 (Ultracontig388), 12q/22q (Ultracontig443), 16q/19q (Ultracontig517) and 22q12/12q24.3 (Ultracontig57; Table 2). On the basis of these data (constrained as they are by the partially complete platypus genome), and the chromosome number estimates presented above, our data suggest that the mammalian ancestral karyotype likely resembled the PAK in terms of chromosome number ($n=23$), and in the majority of the conserved syntenic segmental associations (only 10p/12pq/22qt and 14/15 were not detected in our scans of the platypus and the

former has been regarded as a comparatively weakly supported ancestral chromosome form, see Froenicke *et al.*, 2006). More detailed correspondence between the PAK and the ancestral karyotype for Mammalia is clearly dependent on progress in assembling the platypus genome.

Amniote ancestral configuration

The frog is an appropriate outgroup for defining syntenic segmental associations present in the ancestral amniote karyotype. The *X. tropicalis* genome is estimated at ~1.7 Gbp, distributed over 10 chromosomes or linkage groups (Hellsten *et al.*, 2010). Of this, 769 Mb has been placed onto 691 scaffolds using genetic markers. This paucity of information is further underscored by 200 Mbp being assigned to linkage groups based on inference but without genetic markers (Hellsten *et al.*, 2010), clearly necessitating further experimental studies. Despite this, our scans reveal that most of the ancestral placental syntenic segments are conserved in the frog genome (Figure 2b). In particular, the syntenies 3p/21, 4pq/8p, 7a/16p, 14/15, 12qt/22q and 12pq/22qt are present in some of the *Xenopus* scaffolds; in contrast, there was no evidence of 4q/8p/4pq, 10p/12pq/22qt and 16q/19q (Table 2).

Previous reports have attempted the description of the ancestral amniote genome (Nakatani *et al.*, 2007; Ouangraoua *et al.*, 2009) based on different taxon representation and methodological approaches. Nakatani *et al.*, (2007) defined an ancestral amniote karyotype (AAK) comprising $n=26$. According to the authors, the AAK would present the following ancestral syntenic segmental associations in its chromosomes (see Figure 4 in Nakatani *et al.*, 2007): 12/7b/12/22/7/16/17/22 (AAK 1), Xq/18/8pq/6/5/3p/7/10q (AAK 2), 5 (AAK 3), 2/6/13/3/2 (AAK 4), 6/20 (AAK 5), 10pq (AAK 6), 14/15 (AAK 7), 4 (AAK 8); 1 (AAK 9), 15 (AAK 10), 12/22 (AAK 11), 19/16 (AAK 12), 3/11 (AAK 13), 3/11 (AAK 14), 17 (AAK 15), 17 (AAK 16), 1/16 (AAK 17), 20 (AAK 18), X/5 (AAK 19), 6/19 (AAK 20), 1 (AAK 21), 1 (AAK 22), 8/7/2 (AAK 23), 19p (AAK 24), 11 (AAK 25) and 18/9/5 (AAK 26). Interestingly, the ancestral syntenic segments 10/12/22, 4/8 and 3/21 were not reported in the Nakatani *et al.* (2007) construct. It seems probable that the 4/8 and 3/21 are included in one or more of the unassigned blocks in Nakatani *et al.* (2007) given that they are present in chicken (Robinson and Ruiz-Herrera, 2008) and also frog (present study). More puzzling, however, is 10/12/22, which is not detected in chicken, nor in the frog genome, but is present in opossum and several placental mammals. We therefore view 10/12/22 as a chromosomal signature for Mammalia. Its absence in the platypus genome is due to the low coverage of the assembled sequences or, alternatively, to disruption in the lineage leading to Prototheria.

Tetrapod ancestral configuration

Our comparative genome analyses directed at establishing the likely composition of the tetrapod common ancestor are consistent with those of Kohn *et al.* (2006) with respect to the chromosomal number ($n=18$) and six conserved syntenic segmental associations that it likely contained (that is, 3p/21, 4pq/8p, 7a/16p, 12q/22q, 12pq/22qter and 14/15). We differ with respect to the involvement of 1/19p and 16q/19q suggested in Kohn *et al.* (2006). The inclusion of the former was based on its presence in Afrotheria (Yang *et al.*, 2003). Interestingly this synteny is not found in chicken, platypus, nor in opossum, but 1p manifests as 1p/19p and 1p/19q in different scaffolds of the frog genome. This suggests the existence of 1p, 19p and 19q as separate syntenies in the tetrapod ancestral complement, and their independent assembly in the lineage leading to the frog. The presence of 1p/19q in opossum would therefore represent a convergent change (homoplasy).
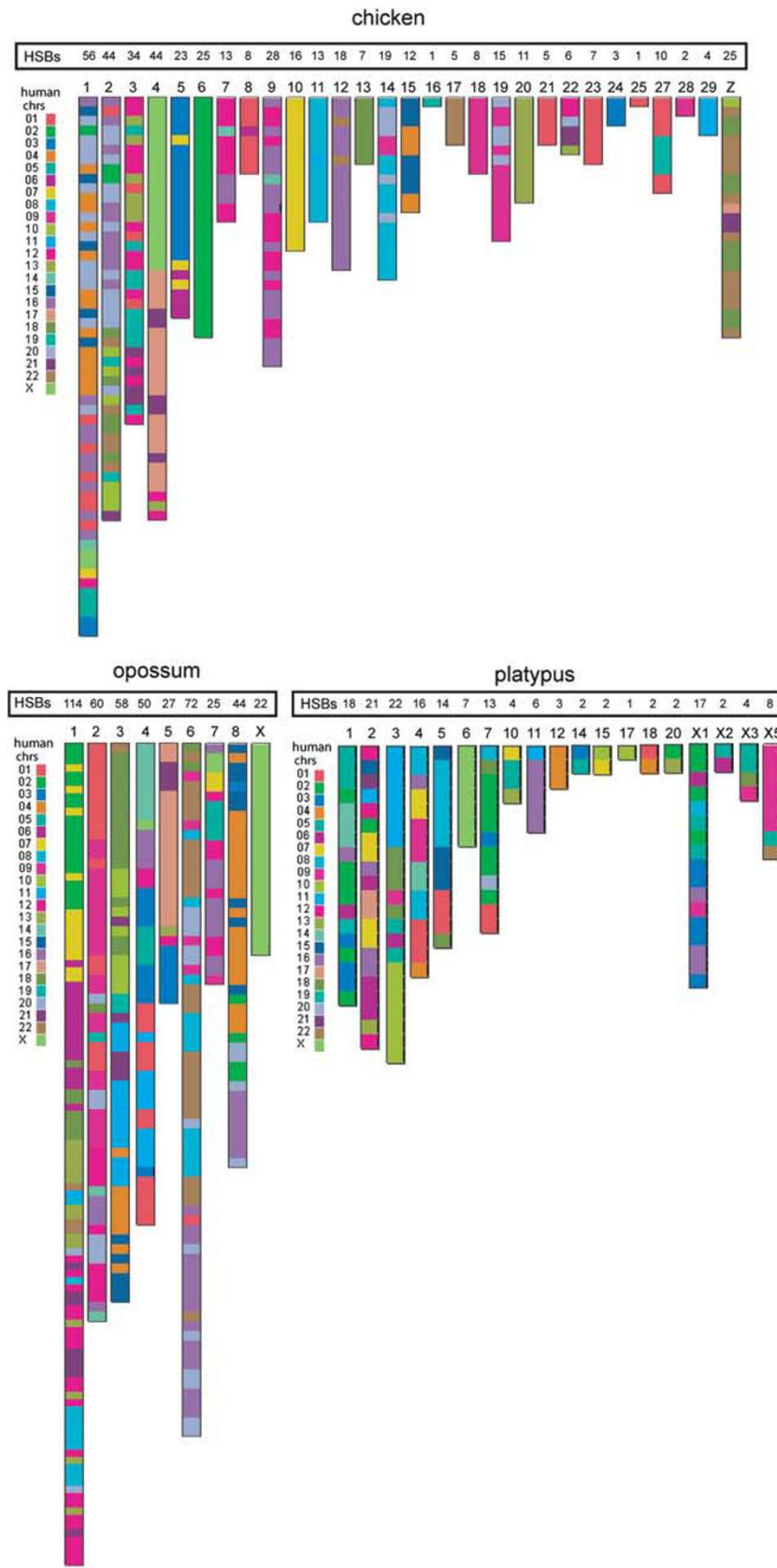
Figure 3 Conserved human chromosomal segments in the genomic assemblies of chicken, opossum and platypus. The human orthologous regions are color-coded and indicated as homologous syntenic blocks (HSBs) in the chromosomes of the respective species. The lengths of the chromosomes are based on homology coverage with the human genome and are not proportional to the chromosomal length.

Feuk L, MacDonald JR, Tang T, Carson AR, Li M, Rao G et al. (2005). Discovery of human inversion polymorphisms by comparative analysis of human and chimpanzee DNA sequence assemblies. *PLoS Genet* **1**: e56.

Froenicke L (2005). Origins of primate chromosomes - as delineated by Zoo-FISH and alignments of human and mouse draft genome sequences. *Cytogenet Genome Res* **108**: 122–138.

Froenicke L, Caldes MG, Graphodatsky A, Muller S, Lyons LA, Robinson TJ et al. (2006). Are molecular cytogenetics and bioinformatics suggesting diverging models of ancestral mammalian genomes? *Genome Res* **16**: 306–310.

Froenicke L, Wienberg J, Stone G, Adams L, Stanyon R (2003). Towards the delineation of the ancestral eutherian genome organization: comparative genome maps of the human and the African elephant (*Loxodonta africana*) generated by chromosome painting. *Proc R Soc Lond B* **270**: 1331–1340.

Glas R, Marshall Graves JA, Toder R, Ferguson-Smith M, O'Brien PC (1999). Cross-species chromosome painting between human and marsupial directly demonstrates the ancient region of the mammalian X. *Mamm Genome* **10**: 1115–1116.

Gregory TR (2011). Animal Genome Size Database. http://www.genomesize.com.

Griffin DK, Robertson LB, Tempest HG, Skinner BM (2007). The evolution of the avian genome as revealed by comparative molecular cytogenetics. *Cytogenet Genome Res* **117**: 64–77.

Hallström BM, Janke A (2010). Mammalian evolution may not be strictly bifurcating. *Mol Biol Evol* **27**: 2804–2816.

Haussler D, O'Brien SJ, Ryder OA, Barker FK, Clamp M, Crawford AJ, Hanner R, Hanotte O, Johnson WE, McGuire JA et al. (2009). Genome 10K: a proposal to obtain whole-genome sequence for 10 000 vertebrate species. *J Hered* **100**: 659–674.

Hayman DL (1990). Marsupial cytogenetics. *Aust J Zool* **37**: 331–349.

Hayman DL, Martin PG (1969). Cytogenetics of marsupials. In: Benirschke K (ed.). *Comparative Mammalian Cytogenetics*. Springer-Verlag: New York.

Hellsten U, Harland RM, Gilchrist MJ, Hendrix D, Jurka J, Kapitonov V et al. (2010). The genome of the Western clawed frog *Xenopus tropicalis*. *Science* **328**: 633–636.

Kehrer-Sawatzki H, Cooper DN (2008). Molecular mechanisms of chromosomal rearrangement during primate evolution. *Chromosome Res* **16**: 41–56.

Kehrer-Sawatzki H, Sandig CA, Goidts V, Hameister H (2005). Breakpoint analysis of the pericentric inversion between chimpanzee chromosome 10 and the homologous chromosome 12 in humans. *Cytogenet Genome Res* **108**: 91–97.

Kirkpatrick M (2010). How and why chromosome inversions evolve. *PLoS Biol* **8**: e1000501.

Kirkpatrick M, Barton N (2006). Chromosome inversions, local adaptation and speciation. *Genetics* **173**: 419–434.

Kohn M, Hogel J, Vogel W, Minich P, Kehrer-Sawatzki H, Graves JA et al. (2006). Reconstruction of a 450-My-old ancestral vertebrate protokaryotype. *Trends Genet* **22**: 203–210.

Kriegs JO, Churakov G, Kiefmann M, Jordan U, Brosius J, Schmitz J (2006). Retroposed elements as archives for the evolutionary history of placental mammals. *PLoS Biol* **4**: e91.

Lee J, Han K, Meyer TJ, Kim HS, Batzer MA (2008). Chromosomal inversions between human and chimpanzee lineages caused by retrotransposons. *PLoS One* **3**: e4047.

Lin CH, Zhao H, Lowcay SH, Shahab A, Bourque G (2010). webMGR: an online tool for the multiple genome rearrangement problem. *Bioinformatics* **26**: 408–410.

Longo MS, Carone DM, , Green ED, O'Neill MJ, O'Neill RJ (2009). Distinct retroelement classes define evolutionary breakpoints demarcating sites of evolutionary novelty. *BMC Genomics* **10**: 334.

Ma J, Zhang L, Suh BB, Raney BJ, Burhans RC, Kent WJ et al. (2006). Reconstructing contiguous regions of an ancestral genome. *Genome Res* **16**: 1557–1565.

Matthey R (1945). L'evoltion de la formule chromosomiale chez les Vertebres. *Experentia* **1**: 78–86.

Matthey R (1958). Chromosomes & systematic position of various African Murinae (Mammalia, Rodentia). *Acta Trop* **15**: 97–117.

Murphy W, Froenicke L, O'Brien SJ, Stanyon R (2003). The origin of human chromosome 1 and its homologues in placental mammals. *Genome Res* **13**: 1880–1888.

Murphy WJ, Eizirik E, Johnson WE, Zhang YP, Ryder OA, O'Brien SJ (2001b). Molecular phylogenetics and the origins of placental mammals. *Nature* **409**: 614–618.

Murphy WJ, Eizirik E, O'Brien SJ, Madsen O, Scally M, Douady CJ et al. (2001a). Resolution of the early placental mammalian radiation using Bayesian phylogenetics. *Science* **294**: 2348–2351.

Murphy WJ, Larkin DM, Everts-van der Wind A, Bourque G, Tesler G, Auvil L et al. (2005). Dynamics of mammalian chromosome evolution inferred from multispecies comparative maps. *Science* **309**: 613–617.

Nakatani Y, Takeda H, Kohara Y, Morishita S (2007). Reconstruction of the vertebrate ancestral genome reveals dynamic genome reorganization in early vertebrates. *Genome Res* **17**: 1254–1265.

Nanda I, Benisch P, Fetting D, Haaf T, Schmid M (2011). Synteny conservation of chicken macrochromosomes 1–10 in different avian lineages revealed by cross-species chromosome painting. *Cytogenet Genome Res* **132**: 165–181.

Naruse K, Hori H, Shimizu N, Kohara Y, Takeda H (2004). Medaka genomics: a bridge between mutant phenotype and gene function. *Mech Dev* **121**: 619–628.

Nie W, O'Brien PC, Ng BL, Fu B, Volobouev V, Carter NP et al. (2009). Avian comparative genomics: reciprocal chromosome painting between domestic chicken (*Gallus gallus*) and the stone curlew (*Burhinus oedicnemus*, Charadriiformes an atypical species with low diploid number. *Chromosome Res* **17**: 99–113.

Nilsson MA, Churakov G, Sommer M, Tran NV, Zemann A, Brosius J et al. (2010). Tracking marsupial evolution using archaic genomic retroposon insertions. *PLoS Biol* **8**: e1000436.

Nishihara H, Satta Y, Nikaido M, Thewissen JG, Stanhope MJ, Okada N (2005). A retroposon analysis of Afrotherian phylogeny. *Mol Biol Evol* **22**: 1823–1833.

O'Brien SJ, Menninger JC, Nash WG (2006). *An Atlas of Mammalian Chromosomes*. John Wiley & Sons, Inc: Hoboken, NJ.

O'Neill RJ, Eldridge MD, Metcalfe CJ (2004). Centromere dynamics and chromosome evolution in marsupials. *J Heredity* **95**: 375–381.

O'Neill RJW, O'Neill MJ, Graves JAM (1998). Undermethylation associated with retro-element activation and chromosome remodeling in an interspecific mammalian hybrid. *Nature* **393**: 68–72.

Ouangraoua A, Boyer F, McPherson A, Tannier E, Chauve C (2009). Prediction of contiguous regions in the amniote ancestral genome. *Lect Notes Comput Sci* **5542**: 173–185.

Peng Z, Elango N, Wildman DE, Yi SV (2009). Primate phylogenomics: developing numerous nuclear non-coding, non-repetitive markers for ecological and phylogenetic applications and analysis of evolutionary rate variation. *BMC Genomics* **10**: 247.

Pham SK, Pevzner PA (2010). DRIMM-Synteny: decomposing genomes into evolutionary conserved segments. *Bioinformatics* **26**: 2509–2516.

Postlethwait JH, Woods IG, Ngo-Hazelett P, Yan YL, Kelly PD, Chu F et al. (2000). Zebrafish comparative genomics and the origins of vertebrate chromosomes. *Genome Res* **10**: 1890–1902.

Rens W, Fu B, O'Brien PC, Ferguson-Smith M (2006). Cross-species chromosome painting. *Nat Protoc* **1**: 783–790.

Rens W, O'Brien PC, Yang F, Solanky N, Perelman P, Graphodatsky AS et al. (2001). Karyotype relationships between distantly related marsupials from South America and Australia. *Chromosome Res* **9**: 301–308.

Richard F, Lombard M, Dutrillaux B (2003). Reconstruction of the ancestral karyotype of eutherian mammals. *Chromosome Res* **11**: 605–618.

Rieseberg LH (2001). Chromosomal rearrangements and speciation. *Trends Ecol Evol* **16**: 351–358.

Robinson TJ, Fu B, Ferguson-Smith MA, Yang F (2004). Cross-species chromosome painting in the golden mole and elephant-shrew: support for the mammalian clades Afrotheria and Afroinsectiphillia but not Afroinsectivora. *Proc Biol Sci* **271**: 1477–1484.

Robinson TJ, Ruiz-Herrera A (2008). Defining the ancestral eutherian karyotype: a cladistic interpretation of chromosome painting and genome sequence assembly data. *Chromosome Res* **16**: 1133–1141.

Robinson TJ, Ruiz-Herrera A, Froenicke L (2006). Dissecting the mammalian genome - new insights into chromosomal evolution. *Trends Genet* **22**: 297–301.

Robinson TJ, Seiffert E (2004). Afrotherian origins and interrelationships: new views and future prospects. *Curr Top Dev Biol* **63**: 37–60.

Ruiz-Herrera A, Castresana J, Robinson TJ (2006). Is mammalian chromosomal evolution driven by regions of genome fragility? *Genome Biol* **7**: R115.

Ruiz-Herrera A, Robinson TJ (2007). Chromosomal instability in Afrotheria: fragile sites, evolutionary breakpoints and phylogenetic inference from genome sequence assemblies. *BMC Evol Biol* **7**: 199.

Svartman M, Stone G, Page JE, Stanyon R (2004). A chromosome painting test of the basal eutherian karyotype. *Chromosome Res* **12**: 45–53.

Svartman M, Stone G, Stanyon R (2006). The ancestral eutherian karyotype is present in Xenarthra. *PLoS Genet* **2**: e109.

Voss SR, Kump DK, Putta S, Pauly N, Reynolds A, Henry R et al. (2011). Origin of amphibian and avian chromosomes by fission, fusion, and retention of ancestral chromosomes. *Genome Res* **21**: 1306–1312.

Waddell PJ, Okada N, Hasegawa M (1999). Towards resolving the interordinal relationships of placental mammals. *Syst Biol* **48**: 1–5.

Waters PD, Dobigny G, Waddell PJ, Robinson TJ (2007). Evolutionary history of LINE-1 in the major clades of placental mammals. *PLoS One* **2**: e158.

Westerman M, Meredith RW, Springer MS (2010). Cytogenetics meets phylogenetics: a review of karyotype evolution in diprotodontian marsupials. *J Hered* **101**: 690–702.

White MJD (1973). *Animal Cytology and Evolution*, 3rd edn. Cambridge University Press: London.

Wienberg J (2004). The evolution of eutherian chromosomes. *Curr Opin Genet Dev* **14**: 657–666.

Woods IG, Wilson C, Friedlander B, Chang P, Reyes DK, Nix R et al. (2005). The zebrafish gene map defines ancestral vertebrate chromosomes. *Genome Res* **15**: 1307–1314.

Yang F, Alkalaeva EZ, Perelman PL, Pardini AT, Harrison WR, O'Brien PC et al. (2003). Reciprocal chromosome painting among human, aardvark, and elephant (superorder Afrotheria) reveals the likely eutherian ancestral karyotype. *Proc Natl Acad Sci USA* **100**: 1062–1066.

Zhao H, Bourque G (2009). Recovering genome rearrangements in the mammalian phylogeny. *Genome Res* **19**: 934–942.

Supplementary Information accompanies the paper on Heredity website (http://www.nature.com/hdy)